Research

Genomic structure of the gene for mouse germ cell nuclear factor (GCNF)

Ute Süsens and Uwe Borgmeyer

Address: Zentrum für Molekulare Neurobiologie, Universität Hamburg, Martinistrasse 52, D-20246 Hamburg, Germany.

Correspondence: Uwe Borgmeyer. E-mail: uwe.borgmeyer@zmnh.uni-hamburg.de

Published: 13 September 2000

Genome ***Biology*** 2000, **1(3)**:research0006.1–0006.3

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2000/1/3/research/0006

Received: 13 July 2000
Accepted: 2 August 2000

Abstract

Background: The germ cell nuclear factor (GCNF, also known as retinoid acid receptor-related testis-associated receptor, neuronal cell nuclear receptor or NR6A1) is an orphan receptor in the nuclear receptor superfamily found in mammals, amphibians and fish. The mouse *Gcnf* gene is expressed in the placenta and the developing nervous system and germ cells, and responds to retinoic acid.

Results: We have defined the intron-exon structure of the mouse *Gcnf* gene and found that it contains 11 exons. Exons 1-4 encode the 75 amino acid amino-terminal domain and exon 4 also encodes the core DNA-binding domain. The carboxy-terminal extension is encoded by exon 5, exons 6 and 7 encode the hinge region, and exons 7-11 encode the putative ligand-binding domain. Unusually, the two zinc-finger motifs in the DNA-binding domain are encoded by separate exons.

Conclusions: The protein-coding region of GCNF is contained in 11 exons. The genomic structure of this nuclear receptor gene will be useful for further studies.

Background

The germ cell nuclear factor (GCNF, NR6A1) is a member of the nuclear receptor superfamily [1,2]. Originally isolated from mouse cDNA libraries, homologs of GCNF have been identified in humans, frogs and fish [3-6]. As no ligand has been identified, GCNF is designated an orphan receptor. Also known as RTR (retinoid acid receptor-related testis-associated receptor) or NCNF (neuronal cell nuclear receptor), evolutionary studies have defined GCNF as the only known member of a sixth subfamily of nuclear receptors [7-9]. The mouse *Gcnf* gene is highly expressed in the developing nervous system, in the labyrinthine layer of the placenta and in the developing germ cells [8,10-12]. Two transcripts of approximately 7.5 kb and 2.4 kb are present in testis, but only the larger transcript is found in somatic cells. Hybridization experiments reveal that the size difference is at least partially due to the use of different polyadenylation sites [13]. Interestingly, GCNF expression is transiently up-regulated and later down-regulated again when embryonal carcinoma cells are triggered to differentiate by retinoic acid [14-16].

Results and discussion

We have isolated genomic clones encompassing the mouse *Gcnf* gene, and have defined the intron-exon structure of the gene. Sequence analysis reveals that the coding region of *Gcnf* comprises 11 exons and 10 introns (Table 1). A bacteriophage lambda library and a cosmid library of genomic DNA of the mouse 129 strain were screened with the full-length *Gcnf* cDNA. The DNA from colonies that hybridized was cloned into pBluescript (SK) for further sequence analysis. Exons 3 and 4 were identified from bacteriophage subclones,

Table 1

Organization of the mouse *Gcnf* gene.

Exon number	Exon size (bp)	cDNA position*	5′ splice donor	Intron number	Intron size (kbp)	3′ splice acceptor
1	>344	1-344	CCGCGCAACGgtgggta	1	ND	ctattgttctctctttagGTTTCT
2	42	345-386	CCAGGCACTAgtaagttc	2	>12	gttctttttgctttgcagATGGAG
3	45	387-431	CATATACCTGgtaagtgg	3	ND	tgacttatccatgtttagTTTCCG
4	243	432-674	AACAGGAAGGgtgagttg	4	>12	gtctacatttccttctagCTATCA
5	56	675-730	ACCAGTCCAGgtgagtcc	5	ND	atccatttcttgccaaagATATCA
6	155	731-885	TATCATCCAGgtgagcta	6	ND	tgaagtttttctctccagTAGGTC
7	228	886-1113	TTGAAGATGGgtgagtta	7	1.238	tcctgtccctgccccagGTATGC
8	255	1114-1368	AACTCCACAGgtgagagc	8	ND	cctgtatctgttctccagATTTAG
9	122	1369-1490	CTGAATCAAGgtgagtag	9	1.408	ttttgtttttgttttcagATATCA
10	153	1491-1643	TACATCGCAGgtaatatt	10	1.567	tctcttccctttacctagGCAAGA
11	>869	1644				

Lower-case letters are used for the intron sequence and capital letters for the exon sequence. The GenBank accession numbers for the exons and the flanking sequences are AF254575S1-AF254575S8. *Relative to GenBank entry MMU09563.

and exons 6-11 were identified in cosmid-derived subclones. Additional intron-exon boundaries and the 5′-untranslated region (5′-UTR) were identified by genome walking analysis following the manufacturer's instructions (Clontech). DNA sequencing was performed on an ABI 377-sequencer using the dye terminator protocol (Perkin Elmer) and on a DNA sequencer model 400 (Li-Cor). The DNA sequences were processed using the Wisconsin Package Version 10.0 of the Genetics Computer Group (GCG), Madison, Wisconsin.

All intron-exon junctions obeyed the GT/AG rule ([17] and Table 1). The location of the intron-exon junctions relative to the peptide sequence is shown in Figure 1. The translational start and stop codons are on exons 1 and 11, respectively. Exon 1 contains the 244 bp untranslated sequence at the 5′ end of the cDNA and codes for the first 33 amino acids (Figure 2). This cDNA, isolated by Hirose *et al.* ([7]; GenBank entry MMU09563), starts with an *Eco*RI site that is present

```
MERDERPPSG GGGGGGSAGF LEPPAALPPP PRNTGFCQDE LAELDPGTTN GE
TDSLTLGQGH IPTVSVPDDRA EQRTCLICGD RATGLHYGII SCEGCKGFFK
RSICNKRVYR CSRDKNCVMS RKQRNRCQYC RLLKCLQMGM NRKTAIREDGM
PGGRNKSIGP VQTISEEEEIER IMSGQEFEEE ANHWSNHGDS DHSSPGNRAS
ESNQPSPGST LSSSTRSVELN GFMAFRDQYM GMSVPPHYQY IPHLFSYSGH
SPLLPPQARS LDPQSYSLIH QLMSAEDLEP LGTPMLIEDGTYAVTQAELFA
LLCRLADELL FRQIAWIKKL PFFCELSIKD YTCLLSSTWQ ELILLSSLTV
YSKQIFGELA DVTAKYSPSD EELHRTFSDEG MEVIERLRIYL YHKFHQLKVS
NEEYACMKAI NFLNQTDIRGL TSASQLEQLN KRYWYICQDF TEYKYTHQPN
RFPDLMMCLP EIRYIATGKMV NVPLEQLPLL FKVVLHSCKT STVKE
```

Figure 1
The location of the different exons in the GCNF amino-acid sequence. The core DNA-binding domain is underlined.

```
TTGGGTTCTC CCTACTTAGG TCTTCCTGTT TTTTTTCCAT CACCCCTTTA TTTGGTAGAG
TCCCGTGTGG GCAGCCTCGT TGGGAGGACT ACATTTCCCA GAATTCCTCA CGGGCATGTG
CGTGGCAGCG GCGCGTGACG TCAGAGGAGG GAGCTGGCCA GTGCTGAGGG GGCGCGGCGC
GGAGGGGCGC GGAGCCGGGC GGCTCAGGGG CCCAGAGAGT GCGGCGGCCG AGAGCCTGCC
GGCCCCTGAC AGCCCCCTCC CCCCGTGGAA GACCAGGACG ACGACTACGA AGGCGCAAGT
CATGGCGGAG CAGCGAAGCC CGAGAGGGCC CTGAGCACCG CCGCATGGAG CGGGACGAAC
GGCCACCTAG CGGAGGGGGA GGCGGCGGGG GCTCGGCGGG GTTCCTGGAG CCGCCCGCCG
CGCTCCCTCC GCCGCCGCGC AACG
```

Figure 2
Sequence of exon 1 of *Gcnf*. The location of the *Eco*RI site (GAATTC) marking the 5′-end of the *Gcnf* cDNA (GenBank entry MMU09563) and the putative translational start codon (ATG) are underlined.

in the genomic DNA. The T at position 174 is a G in our genomic isolate, which could represent a genomic variant. As no promoter has been identified for *Gcnf*, the sequence preceding the *Eco*RI site may contain promoter elements. It is also possible, however, that the promoter precedes a not-yet-identified additional exon in the 5′-UTR of *Gcnf*.

The amino-terminal domain of 75 amino acids is encoded by exons 1-4. Exon 4 also codes for the core DNA-binding domain (DBD) of 66 amino acids and for three additional amino acids (Figure 1). The DBD consists of two zinc-finger motifs that are encoded by separate exons in most vertebrate nuclear receptor genes, except for those of the COUP transcription factor subfamily. Evolutionary studies do not provide further evidence that these receptors are closely related to GCNF. A further domain important for DNA binding and for homodimeric interactions, and known as the DBD carboxy-terminal extension, is encoded by the 56 bp of exon 5. The sizes of intron 2 and intron 4 were determined by

PCR amplification of mouse genomic DNA. Exons 6 and 7 code for the hinge region, whereas exons 7-11 code for the putative ligand-binding domain. A variant of the typical AUAAA polyadenylation signal (AGUAAA) and the cleavage site that is used in the testis are part of the eleventh exon [13].

Conclusions

The protein-coding region of GCNF is contained in 11 exons. Additional studies will be required to define the regulatory/promoter region. We think the genomic structure of this first, and at present only, member of the sixth subfamily of nuclear receptors will be useful for further studies of this unique receptor.

Acknowledgements

We would like to thank H.C. Schaller for her support of this work in her institute, and M. Däumigen and W. Kullmann for sequencing.

References

1. Greschik H, Schüle R: **Germ cell nuclear factor: an orphan receptor with unexpected properties.** *J Mol Med* 1998, **76:**800-810.
2. Cooney AJ, Katz D, Hummelke GC, Jackson KJ: **Germ cell nuclear factor: an orphan receptor in search of a function.** *Am Zool* 1999, **39:**796-806.
3. Chen F, Cooney AJ, Wang Y, Law SW, O'Malley BW: **Cloning of a novel orphan receptor (GCNF) expressed during germ cell development.** *Mol Endocrinol* 1994, **8:**1434-1444.
4. Süsens U, Borgmeyer U: **Characterization of the human germ cell nuclear factor gene.** *Biochim Biophys Acta* 1996, **1309:**179-182.
5. Joos T, David R, Dreyer C: **xGCNF, a nuclear orphan receptor is expressed during neurulation in *Xenopus laevis*.** *Mech Dev* 1996, **60:**45-57.
6. Braat AK, Zandbergen MA, de Vries E, van der Burg B, Bogerd J, Goos HJ: **Cloning and expression of the zebrafish germ cell nuclear factor.** *Mol Reprod Dev* 1999, **53:**369-375.
7. Hirose T, O'Brien DA, Jetten AM: **RTR: a new member of the nuclear receptor superfamily that is highly expressed in murine testis.** *Gene* 1995, **152:**247-251.
8. Bauer U-M, Schneider-Hirsch S, Reinhardt S, Pauly T, Maus A, Wang F, Heiermann R, Rentrop M, Maelicke A: **Neuronal cell nuclear factor - a nuclear receptor possibly involved in the control of neurogenesis and neuronal differentiation.** *Eur J Biochem* 1997, **249:**826-837.
9. Laudet V: **Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor.** *J Mol Endocrinol* 1997, **19:**207-226.
10. Süsens U, Aguiluz JB, Evans RM, Borgmeyer U: **The germ cell nuclear factor mGCNF is expressed in the developing nervous system.** *Dev Neurosci* 1997, **19:**410-420.
11. Morasso MI, Grinberg A, Robinson G, Sargent TD, Mahon KA: **Placental failure in mice lacking the homeobox gene Dlx3.** *Proc Natl Acad Sci USA* 1999, **96:**162-167.
12. Katz D, Niederberger C, Slaughter GR, Cooney AJ: **Characterization of germ cell-specific expression of the orphan nuclear receptor, germ cell nuclear factor.** *Endocrinology* 1997, **138:**4364-4372.
13. Zhang YL, Akmal KM, Tsuruta JK, Shang,Q, Hirose T, Jetten AM, Kim KH, O'Brien DA: **Expression of germ cell nuclear factor (GCNF/RTR) during spermatogenesis.** *Mol Reprod Dev* 1998, **50:**93-102.
14. Lei W, Hirose T, Zhang L-X, Adachi H, Spinella MJ, Dmitrovsky E, Jetten AM: **Cloning of the human orphan receptor germ cell nuclear factor/retinoid receptor-related testis-associated receptor and its differential regulation during embryonal carcinoma cell differentiation.** *J Mol Endocrinol* 1997, **18:**167-176.
15. Heinzer C, Süsens U, Schmitz TP, Borgmeyer U: **Retinoids induce differential expression and DNA binding of the mouse germ cell nuclear factor.** *Biol Chem* 1998, **379:**349-359.
16. Schmitz TP, Süsens U, Borgmeyer U: **DNA binding, protein interaction and differential expression of the human germ cell nuclear factor.** *Biochim Biophys Acta* 1999, **1446:**173-180.
17. Shapiro MB, Senapathy P: **RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression.** *Nucleic Acids Res* 1987, **15:**7155-7174.